

# Parallel Direct Solution of the Ensemble Square Root Kalman Filter Equations with Observation Principal Components

JEFFREY L. STEWARD

*University of California, Los Angeles, Los Angeles, California*

ALTUĞ AKSOY

*University of Miami, Miami, Florida*

ZIAD S. HADDAD

*University of California, Los Angeles, Los Angeles, California*

(Manuscript received 15 July 2016, in final form 20 June 2017)

## ABSTRACT

The ensemble square root Kalman filter (ESRF) is a variant of the ensemble Kalman filter used with deterministic observations that includes a matrix square root to account for the uncertainty of the unperturbed ensemble observations. Because of the difficulties in solving this equation, a serial approach is often used where observations are assimilated sequentially one after another. As previously demonstrated, in implementations to date the serial approach for the ESRF is suboptimal when used in conjunction with covariance localization, as the Schur product used in the localization does not commute with assimilation. In this work, a new algorithm is presented for the direct solution of the ESRF equations based on finding the eigenvalues and eigenvectors of a sparse, square, and symmetric positive semidefinite matrix with dimensions of the number of observations to be assimilated. This is amenable to direct computation using dedicated, massively parallel, and mature libraries. These libraries make it relatively simple to assemble and compute the observation principal components and to solve the ESRF without using the serial approach. They also provide the eigenspectrum of the forward observation covariance matrix. The parallel direct approach described in this paper neglects the near-zero eigenvalues, which regularizes the ESRF problem. Numerical results show this approach is a highly scalable parallel method.

## 1. Introduction

Modern data assimilation methods for numerical weather prediction are called upon to merge up to the order of  $10^6 - 10^7$  observations (Lahoz and Schneider 2014) of the surface and atmosphere with model forecasts consisting of the order of  $10^8$  or more state variables to improve initial conditions and to produce probability estimates. As the size of this problem continues to increase, especially as remotely sensed satellite data become increasingly prevalent, it is essential to develop scalable methods that can handle increasing problem sizes in efficient and accurate ways.

In this paper we will explore a new method for directly solving the square root ensemble Kalman filter (EnKF) equations without using a serial approximation or perturbed observations. This implementation is scalable across many processing elements and takes advantage of recent improvements in both theoretical and computational linear algebra. It avoids a major problem that occurs with the serial approximation to the square root EnKF when used in the presence of covariance localization—namely, that the ordering of the observations becomes important, since the covariance localization operation does not commute (Nerger 2015; Bishop et al. 2015). While Bishop et al. (2015) introduced the consistent hybrid ensemble filter (CHEF), which provides consistent serial observation processing in a non-square root EnKF framework by using perturbed observations, this manuscript

---

*Corresponding author:* Jeffrey L. Steward, [jsteward@jifresse.ucla.edu](mailto:jsteward@jifresse.ucla.edu)

addresses how to make the square root EnKF equations consistent with covariance localization without the use of perturbed observations using a parallel direct approach.

The EnKF was originally introduced by Evensen (1994) as a method to estimate the covariances used in the Kalman filter equation (Kalman 1960) using an ensemble of model predictions. This innovation made it feasible to run statistical data assimilation problems even on very large dynamical systems of size  $N_{\text{state}}$ , where the full covariance matrices of size  $(N_{\text{state}} \times N_{\text{state}})$  would be prohibitively large to store and invert. Instead, a small ensemble of size  $(N_{\text{state}} \times N_{\text{ens}})$ , where  $N_{\text{ens}} \ll N_{\text{state}}$ , is used to estimate the necessary covariance matrices.

Two crucial issues arise in the use of the EnKF as originally formulated. The first crucial issue is the systematic underestimation of ensemble spread that occurs when assimilating the same observations to update both the ensemble mean and the covariance with the unmodified Kalman equations. Two main families of solutions have been devised to address this problem. The first approach (Houtekamer and Mitchell 1998; Burgers et al. 1998) is to perturb observations with independently sampled noise to update each ensemble member. However, this introduces sampling error, which causes the filter to be suboptimal—especially when  $N_{\text{ens}}$  is small—as demonstrated by Whitaker and Hamill (2002). The other approach, which uses the same nonperturbed observation for each ensemble member, is the ensemble square root filter (ESRF). This method explicitly accounts for the underrepresentation of error stemming from the use of deterministic observations by adding a square root term to the Kalman update for the ensemble. Various flavors of ESRF have been developed (Bishop et al. 2001; Anderson 2001; Whitaker and Hamill 2002) that Tippett et al. (2003) showed are all equivalent, in the sense they perform analysis in the same vector space and find the same covariance.

The second crucial issue is that the sample covariance, based on a low-rank estimate, may contain spurious correlations between two distant points. This issue decreases in severity as the ensemble size grows, but because by design  $N_{\text{ens}} \ll N_{\text{state}}$ , the issue of spurious correlations cannot be completely eliminated by increasing the ensemble size. Therefore, a covariance localization method, which uses a Schur product (component-wise multiplication) to zero out correlations further than a specified distance (Gaspari and Cohn 1999; Houtekamer and Mitchell 2001; Hamill et al. 2001), is typically employed to reduce spurious

correlations. This also increases the rank of the covariance matrices.

Without covariance localization, the ESRF equations can be solved in a serial fashion as the Kalman gain commutes. The commutativity property of the serial ESRF solution is lost when covariance localization is used, as the Schur product is a nonlinear operation, and the serial ESRF approximation does not correctly update the high-rank localized matrix. Issues began to appear in the combination of the serial ESRF with covariance localization in the differences in results in Holland and Wang (2013) using a simplified primitive equations model and, subsequently, Nerger (2015) and Bishop et al. (2015) clearly demonstrated the suboptimality of the serial approach when used without perturbed observations, while Bishop et al. (2015) presented the CHEF framework that provides consistent updates to the mean and covariances with perturbed observations. In the presence of covariance localization without perturbed observations, assimilating all observations at once is a theoretically preferable method. However, the challenge of how to implement this in a scalable way in actual full-scale data assimilation systems seems daunting at first. In this paper, we demonstrate how to take advantage of modern, massively parallel linear algebra system libraries to solve these equations in a scalable way. This becomes the basis for our parallel direct ESRF filter.

#### a. Motivation

Several parallel EnKF implementations are described in the literature, including Keppenne and Rienecker (2002), Zhang et al. (2005), Anderson and Collins (2007), and Wang et al. (2013). These methods do not specifically address the issue of the clash between covariance localization and the serial approach that causes the ordering of observations to become important. Furthermore, as shown in Anderson and Collins (2007), an additional complexity that arises when using the serial filter in a parallel setting is that the forward-calculated observations  $H(\mathbf{x})$  (where  $H$  is the possibly nonlinear operator mapping the model space to the observation space and  $\mathbf{x}$  is the model/state variable) must be either recomputed (which would require communication among processors) or treated as part of the augmented state that is also updated during the serial filter. This second approach serves to alleviate a major problem with parallelizing the serial filter, but it still requires observations to be assimilated one at a time. Indeed, the very concept of “serial” observations seems antithetical to parallelization, by definition. Therefore, perhaps a reevaluation of the serial filter is in order, as the data assimilation problem size continues to grow.

In Bishop et al. (2015) a method is described to ensure the assimilation update does not depend upon the order of the observations and serial batches of perturbed observations are assimilated by correctly updating the true high-rank localized covariance matrices. In this paper we present a method for solving the global ESRF equations with covariance localization such that the solution is likewise independent of ordering, but in our case the observations are assimilated all at once. Because we work with eigenpairs of the forward-calculated observation covariance matrix, it is natural for us to examine the spectrum of this matrix. We demonstrate that the effective rank of the forward observation covariance matrix decreases as a function of localization length scale. This method can therefore be used to study the eigenspectrum of different localization techniques.

### b. The HWRf Ensemble Data Assimilation System

For our base data assimilation system, we use the Hurricane Ensemble Data Assimilation System (HEDAS), developed by the Hurricane Research Division (HRD) of the NOAA Atlantic Oceanographic and Meteorological Laboratory (AOML) under NOAA's Hurricane Forecast Improvement Program (HFIP) (Aksoy et al. 2012, 2013; Aksoy 2013; Vukicevic et al. 2013; Aberson et al. 2015). HEDAS is based on a serial implementation of the ESRF and is specifically designed to assimilate hurricane observations collected by reconnaissance aircraft. Its focus is to generate a vortex that reasonably matches flight data, and it includes covariance localization at heuristically determined length scales. The application uses 30 ensemble members. The initial and lateral boundary ensemble perturbations are obtained from operational National Centers for Environmental Prediction Global Ensemble Forecast System analyses. Aksoy et al. (2013) showed that the primary circulation in HEDAS exhibits correlations with their observed counterparts of 87%–97%. HEDAS also produces good analyses in other aspects of the structure of the primary circulation, such as the radius of maximum winds, wave-number 1 asymmetry, and storm size. As noted by Nerger (2015), the suboptimal effect of the non-commutativity of observations and covariance localization is hypothesized to be small in realistic systems when the assimilation produces small changes to the background, and as HEDAS distributes observations among multiple frequently spaced assimilation cycles (Aksoy 2013), this is likely to be the case, especially after the first assimilation cycle. Nonetheless, in this paper we are interested in describing a new scalable, parallel, optimal ESRF core built upon the HEDAS

processing framework that overcomes the limitations described above.

## 2. Direct solution of square root observation filter equations: Theoretical aspects

In this section we develop the theoretical basis for our parallel direct solution to the ESRF equations using covariance localization. The important symbols referenced in this paper are listed in the appendix to assist the reader.

### a. Solution of ESRF matrix function via eigenpairs

For the ensemble mean  $\bar{\mathbf{x}}$  of size  $N_{\text{state}} \times 1$  and ensemble perturbations  $\mathbf{X}'$  of size  $N_{\text{state}} \times N_{\text{ens}}$ , the square root ensemble Kalman filter without perturbed observations (Whitaker and Hamill 2002) can be written as an update to the analysis  $a$  from the previous forecast  $f$  as

$$\begin{aligned}\bar{\mathbf{x}}_a &= \bar{\mathbf{x}}_f + \mathbf{K}(\mathbf{y} - \overline{H(\mathbf{X}_f)}) \\ \mathbf{X}'_a &= \mathbf{X}'_f + \tilde{\mathbf{K}}(\mathbf{0} - H\mathbf{X}),\end{aligned}\quad (1)$$

where  $\mathbf{y}$  ( $N_{\text{obs}} \times 1$ ) is the observations,  $\overline{H(\mathbf{X}_f)}$  ( $N_{\text{obs}} \times 1$ ) is the mean of the forward-calculated observations,  $H\mathbf{X}_{i,j} = h_i(\mathbf{X}_f^{(j)}) - \overline{h_i(\mathbf{X}_f)}$  is the mean-subtracted  $i$ th observation operator acting on the  $j$ th ensemble member  $\mathbf{X}_f^{(j)}$ . Term  $H\mathbf{X}$  is  $N_{\text{obs}} \times N_{\text{ens}}$  (as is  $\mathbf{0}$ , a matrix filled with zeros), and the assumption is that  $N_{\text{state}} \gg N_{\text{obs}} \gg N_{\text{ens}}$ . The traditional Kalman gain,  $\mathbf{K}$  ( $N_{\text{state}} \times N_{\text{obs}}$ ), is

$$\mathbf{K} = \mathbf{C}_{\mathbf{x}, H\mathbf{x}} \mathbf{D}^{-1}, \quad (2)$$

where  $\mathbf{C}_{\mathbf{x}, H\mathbf{x}} = \text{cov}(\mathbf{x}_f, H(\mathbf{x}_f))$  is the covariance between  $\mathbf{x}_f$  (an  $N_{\text{state}} \times 1$  random variable representing the previous forecast) and  $H(\mathbf{x}_f)$  (the observation operator acting on this random variable). Equation  $\mathbf{D} = \mathbf{C}_{H\mathbf{x}, H\mathbf{x}} + \mathbf{R}$  for  $\mathbf{C}_{H\mathbf{x}, H\mathbf{x}} = \text{cov}(H(\mathbf{x}_f), H(\mathbf{x}_f))$ , and  $\mathbf{R}$  is the observation error covariance  $\text{cov}(\mathbf{y}_t - H(\mathbf{x}_f))$  for a random variable  $\mathbf{y}_t$  representing the true observations without observation noise.

Term  $\tilde{\mathbf{K}}$  ( $N_{\text{state}} \times N_{\text{obs}}$ ), the correction from using non-perturbed observations, is

$$\tilde{\mathbf{K}} = \mathbf{C}_{\mathbf{x}, H\mathbf{x}} \mathbf{D}^{-1/2} (\sqrt{\mathbf{D}} + \sqrt{\mathbf{R}})^{-1}. \quad (3)$$

These equations produce the minimum error variance estimate of the state and covariance given the observations, a result that holds even with non-Gaussian errors and nonlinear observations (Kalnay 2003).

As it is difficult to compute the square roots and inverses in Eqs. (1)–(3), the so-called serial approach

is usually taken, but in this work we aim to directly solve Eq. (1). Since  $\tilde{\mathbf{K}}$  involves the square root<sup>1</sup> of  $\mathbf{D}$  and  $\mathbf{R}$ ,  $\tilde{\mathbf{K}}$  is more difficult to work with than  $\mathbf{K}$ . These square roots will require analysis of eigenvalue/eigenvector pairs.

To begin, it simplifies our analysis greatly if we remove the square root of  $\mathbf{R}$  by noting that  $\mathbf{R}$  can be written as the identity matrix  $\mathbf{I}$  through the transformation

$$\mathbf{y} = \mathbf{R}_{\text{old}}^{-1/2} \mathbf{y}_{\text{old}}, \quad (4)$$

where the “old” subscript represents the untransformed observation. In addition, the observation operator is also scaled as  $H(\mathbf{x}) = \mathbf{R}_{\text{old}}^{-1/2} H_{\text{old}}(\mathbf{x})$ . The effect of this transformation is that  $\mathbf{R}$  is now identity. This changes the covariances from the original ESRF problem, but assimilating observations with different units (or no units, as with this transformation) should not lead to drastically different analyses.<sup>2</sup> Note that for the diagonal observation error matrix  $\mathbf{R}_{\text{old}}$  typically considered, multiplying by  $\mathbf{R}_{\text{old}}^{-1/2}$  amounts to dividing each observation by the standard deviation of the observation error; for correlated nondiagonal  $\mathbf{R}_{\text{old}}$ , this transformation removes that correlation by using principal components.

We now begin our analysis. Rearranging Eq. (3), we have

$$\tilde{\mathbf{K}} = \mathbf{C}_{\mathbf{x}, H\mathbf{x}} \mathbf{M}^{-1} \quad (5)$$

for  $\mathbf{M} = (\sqrt{\mathbf{D}} + \sqrt{\mathbf{R}})\sqrt{\mathbf{D}}$ . By Eq. (4),  $\mathbf{M} = \mathbf{D} + \sqrt{\mathbf{D}}$ . Let  $\lambda_i$ ,  $\mathbf{v}_i$  denote the  $i$ th eigenpair of  $\mathbf{C}_{H\mathbf{x}, H\mathbf{x}}$ . Then

$$\mathbf{M}\mathbf{v}_i = \mathbf{C}_{H\mathbf{x}, H\mathbf{x}} \mathbf{v}_i + \mathbf{v}_i + (\mathbf{C}_{H\mathbf{x}, H\mathbf{x}} + \mathbf{I})^{1/2} \mathbf{v}_i. \quad (6)$$

For some square matrix  $\mathbf{T}$ , as the eigenvectors for  $(\mathbf{T} + \alpha\mathbf{I})^p$  are the same as  $\mathbf{T}$  for any scalar  $\alpha$  and  $p$ , while the eigenvalues are  $(\lambda_i + \alpha)^p$  (Higham 2008), we have

$$\mathbf{M}\mathbf{v}_i = \left( \lambda_i + 1 + (\lambda_i + 1)^{1/2} \right) \mathbf{v}_i; \quad (7)$$

therefore,

$$\mathbf{M}^{-1} \mathbf{v}_i = \lambda'_i \mathbf{v}_i \quad (8)$$

for

$$\lambda'_i = \frac{1}{\lambda_i + 1 + (\lambda_i + 1)^{1/2}}. \quad (9)$$

Note that  $\lambda_i \geq 0$  as  $\mathbf{C}_{H\mathbf{x}, H\mathbf{x}}$  is a symmetric positive semidefinite covariance matrix. As  $\lambda_i \rightarrow \infty$ ,  $\lambda'_i$  goes to 0, while for  $\lambda_i \rightarrow 0$ ,  $\lambda'_i$  goes to 1/2.

With covariance localization, without the loss of symmetry or positive semidefiniteness, the matrix  $\mathbf{C}_{H\mathbf{x}, H\mathbf{x}}$  in our implementation is formed by a Schur product,

$$\mathbf{C}_{H\mathbf{x}, H\mathbf{x}} = \rho_{\mathbf{y}, \mathbf{y}} \circ \mathbf{Q}_{H\mathbf{x}, H\mathbf{x}}, \quad (10)$$

where  $\rho_{\mathbf{y}, \mathbf{y}}$  is the localization matrix arising from a localization function (Gaspari and Cohn 1999)  $\ell$  such that

$$(\rho_{\mathbf{y}, \mathbf{y}})_{ij} = \ell(d_{ij} | L_{ij}), \quad (11)$$

where  $d_{ij}$  is the distance between the location of the  $i$ th and  $j$ th observations,<sup>3</sup> and  $L_{ij}$  is the characteristic length scale for the localization function  $\ell$ . Matrix  $\mathbf{Q}_{H\mathbf{x}, H\mathbf{x}}$  is defined as

$$\mathbf{Q}_{H\mathbf{x}, H\mathbf{x}} = \frac{H\mathbf{X}(H\mathbf{X})^T}{N_{\text{ens}} - 1}. \quad (12)$$

By the Schur product rank inequality theorem,

$$\text{rank}(\mathbf{C}_{H\mathbf{x}, H\mathbf{x}}) \leq \text{rank}(\rho_{\mathbf{y}, \mathbf{y}}) \text{rank}(H\mathbf{X}). \quad (13)$$

Since the rank of  $H\mathbf{X}$  is  $N_{\text{ens}}$ , we have

$$\text{rank}(\mathbf{C}_{H\mathbf{x}, H\mathbf{x}}) \leq N_{\text{ens}} \text{rank}(\rho_{\mathbf{y}, \mathbf{y}}). \quad (14)$$

Assume for a moment  $L_{ij} = L$  for all  $i$  and  $j$ . The rank of  $\rho_{\mathbf{y}, \mathbf{y}}$  depends on  $L$  and the distances between observations. When  $L \rightarrow 0$ , all observations become completely independent (as  $\ell_{ij} = \delta_{ij}^j$  for the Kronecker  $\delta$ ) and  $\mathbf{C}_{H\mathbf{x}, H\mathbf{x}}$  is full rank, while when  $L \rightarrow \infty$ , all

<sup>1</sup> In this paper by  $\mathbf{A} = \mathbf{B}^{1/2}$  we mean  $\mathbf{A}\mathbf{A} = \mathbf{B}$  rather than the alternative definition  $\mathbf{A}\mathbf{A}^T = \mathbf{B}$ .

<sup>2</sup> While a similar normalization technique is used in Bishop et al. (2001), our main motivation is to treat the  $\mathbf{R}_{\text{old}}^{-1/2}$  scaling term as a preprocessing step in order to avoid carrying it through the equations. This changes the original assimilation problem in the presence of covariance localization [see Eqs. (10) and (26) below] so that the scaled observations and operators are localized and assimilated. This approach is otherwise equivalent to the nonscaled problem but is easier to solve.

<sup>3</sup> Note to ensure positive semidefiniteness of  $\rho_{\mathbf{y}, \mathbf{y}}$ , the distance defined here must follow the rules of a metric, namely, the distance is nonnegative, a distance is zero if and only if the points are the same, the distance between two points is symmetric, and the triangle inequality between three points holds. As a practical note, the mismatch of different length scales  $L_{ij}$  may violate the third or fourth requirements.

observations are completely correlated (as  $\ell_{ij} = 1$  for all  $i, j$ ) and  $\text{rank}(\mathbf{C}_{H\mathbf{x}, H\mathbf{x}}) = N_{\text{ens}}$ , the same as without localization.

Returning our attention to Eq. (10), for  $\lambda_i \approx 0$ , the corresponding  $\lambda'_i$  of  $\mathbf{M}^{-1}$  will be approximately 1/2 (which makes  $\mathbf{M}$  invertible regardless of whether  $\mathbf{C}_{H\mathbf{x}, H\mathbf{x}}$  is). The remaining eigenvectors corresponding to the approximately 1/2 eigenvalues are orthogonal to  $W = \text{span}(\mathbf{v}_i)$  for  $i = 1, \dots, r_{\text{true}}$ , where  $r_{\text{true}} = \text{rank}(\mathbf{C}_{H\mathbf{x}, H\mathbf{x}})$ .

With this in mind, suppose we set all eigenvalues  $\lambda'_i$  to 1/2 for  $i > r$  for some  $r \leq r_{\text{true}}$  such that  $\lambda_r$  is small. Thus, define the matrix  $\mathbf{M}_r^{-1}$  operating on the  $i$ th eigenvector as

$$\mathbf{M}_r^{-1} \mathbf{v}_i = \begin{cases} \lambda'_i \mathbf{v}_i & \text{if } i \leq r \\ \frac{1}{2} \mathbf{v}_i & \text{otherwise.} \end{cases} \quad (15)$$

The difference between  $\mathbf{M}^{-1} \mathbf{v}_i$  and  $\mathbf{M}_r^{-1} \mathbf{v}_i$  is 0 for  $i \leq r$ , and for  $i > r$  it is

$$(\mathbf{M}^{-1} - \mathbf{M}_r^{-1}) \mathbf{v}_i = \left( \lambda'_i - \frac{1}{2} \right) \mathbf{v}_i \quad (16)$$

by Eq. (7),

$$\lambda'_i - \frac{1}{2} = \frac{1 - \frac{1}{2} (\lambda_i + 1 + (\lambda_i + 1)^{1/2})}{\lambda_i + 1 + (\lambda_i + 1)^{1/2}}. \quad (17)$$

The Taylor series expansion around 0 for  $(\lambda_i + 1)^{1/2} = 1 + \frac{\lambda_i}{2} - \frac{\lambda_i^2}{8} + \dots$ . Substituting,

$$\lambda'_i - \frac{1}{2} = \frac{1 - \frac{1}{2} \left( 2 + \frac{3}{2} \lambda_i - \frac{\lambda_i^2}{8} + \dots \right)}{2 + \frac{3}{2} \lambda_i - \frac{\lambda_i^2}{8} + \dots}. \quad (18)$$

Simplifying,

$$\lambda'_i - \frac{1}{2} = \frac{-3\lambda_i + \frac{\lambda_i^2}{4} + \dots}{8 + 2 \left( 3\lambda_i - \frac{\lambda_i^2}{4} + \dots \right)}. \quad (19)$$

This gives an upper bound of the error as

$$\left| \lambda'_i - \frac{1}{2} \right| \leq \frac{3}{8} \lambda_i \quad (20)$$

for  $0 \leq \lambda_i \leq 1$ . Since the two (spectral) norm of some matrix  $\mathbf{T}$  is the square root of the largest eigenvalue of  $\mathbf{T}^T \mathbf{T}$ , we have

$$\|\mathbf{M}^{-1} - \mathbf{M}_r^{-1}\| \leq \frac{3}{8} \lambda_{r+1}, \quad (21)$$

where  $r + 1$  is the index of the first neglected eigenvalue. Therefore, if we can devise a numerical method to find the largest eigenvalue eigenpairs of the  $N_{\text{obs}} \times N_{\text{obs}}$  sparse matrix  $\mathbf{C}_{H\mathbf{x}, H\mathbf{x}}$ , we can accurately and efficiently directly solve the difficult part of the square root ensemble Kalman filter equations. In particular, as the space spanned by the constant 1/2 eigenvalues is orthogonal to the space spanned by the first  $r$  eigenvectors, we have

$$\mathbf{M}_r^{-1} \mathbf{Z}_j = \sum_{i=1}^r \lambda'_i \alpha_{i,j} \mathbf{v}_i + \frac{1}{2} (\mathbf{Z}_j - \text{proj}_{V_r}(\mathbf{Z}_j)), \quad (22)$$

where  $\mathbf{Z}_j = (\mathbf{0} - H\mathbf{X})_j$  is the  $j$ th negative observation perturbation;  $\alpha_{i,j}$  is the projection of  $\mathbf{Z}_j$  along  $\mathbf{v}_i$ , that is,

$$\alpha_{i,j} = \mathbf{v}_i^T \mathbf{Z}_j; \quad (23)$$

and  $\text{proj}_{V_r}(\mathbf{Z}_j)$  is the projection of  $\mathbf{Z}_j$  into the space  $V_r$  spanned by the first  $r$  eigenvectors, that is,

$$\text{proj}_{V_r}(\mathbf{Z}_j) = \sum_{i=1}^r \alpha_{i,j} \mathbf{v}_i. \quad (24)$$

For  $\mathbf{K}$ , the same method can be used on  $\mathbf{D}$  instead of  $\mathbf{M}$ , that is,

$$\mathbf{D}^{-1} \mathbf{Y} = \sum_{i=1}^r \frac{\beta_i}{\lambda_i + 1} \mathbf{v}_i + \mathbf{Y} - \text{proj}_{V_r}(\mathbf{Y}) \quad (25)$$

for  $\mathbf{Y} = \mathbf{y} - \overline{H(\mathbf{x}_f)}$  and  $\beta_i = \mathbf{v}_i^T \mathbf{Y}$ .

Equation (1) is now solved by premultiplying  $\mathbf{M}^{-1} \mathbf{Z}_j$  and  $\mathbf{D}^{-1} \mathbf{Y}$  by

$$\mathbf{C}_{\mathbf{x}, H\mathbf{x}} = \rho_{\mathbf{x}, \mathbf{y}} \circ \mathbf{Q}_{\mathbf{x}, H\mathbf{x}} \quad (26)$$

for

$$(\rho_{\mathbf{x}, \mathbf{y}})_{i,j} = \ell(d_{i,j} | L_{i,j}), \quad (27)$$

where  $d_{i,j}$  is the distance between the location of the model state  $i$  and observation  $j$  with the same localization function as Eq. (11), and

$$\mathbf{Q}_{\mathbf{x}, H\mathbf{x}} = \frac{\mathbf{X}'_f (H\mathbf{X})^T}{N_{\text{ens}} - 1}. \quad (28)$$

#### b. Rank-deficient observation covariance matrix

In the previous section, we developed an algorithm to solve the ESRF equations directly by finding



the  $r$  largest eigenvalues of  $\mathbf{C}_{H\mathbf{x},H\mathbf{x}}$  and proved that the eigenpair-based method solves Eq. (1) to within a configurable tolerance. However, the terms  $\frac{1}{2}[\mathbf{Z}_i - \text{proj}_{V_r}(\mathbf{Z}_i)]$  and  $\mathbf{Y} - \text{proj}_{V_r}(\mathbf{Y})$  are problematic. If the  $\mathbf{C}_{H\mathbf{x},H\mathbf{x}}$  matrix is full rank, then these terms will be zero, since  $V_r$  will form a basis for  $\mathbb{R}^{N_{\text{obs}}}$  and  $\mathbf{a} = \text{proj}_{V_r}(\mathbf{a})$  for all  $\mathbf{a}$ . In the case that  $\mathbf{C}_{H\mathbf{x},H\mathbf{x}}$  is rank deficient, these terms are allowed to contribute to the inverse directly without input from the ensemble covariance. In other words, if  $\mathbf{C}_{H\mathbf{x},H\mathbf{x}}$  is rank deficient, then there will be a vector  $\mathbf{w}$  such that  $\mathbf{C}_{H\mathbf{x},H\mathbf{x}}\mathbf{w} = 0$  (i.e.,  $\mathbf{w}$  is in the null space of  $\mathbf{C}_{H\mathbf{x},H\mathbf{x}}$ ), and therefore this portion of  $\mathbf{z}_w = \mathbf{z} + \mathbf{w}$  can be added into the inverse without being impacted by the observation covariance. This  $\mathbf{w}$  term represents pure noise and degrades the quality of the solution.

To see this more clearly, consider that the eigenvalues  $\lambda_i$  will approach zero until  $i = r_{\text{true}}$ , but there is a fundamental difference between eigenpairs for  $i \leq r_{\text{true}}$  and  $i > r_{\text{true}}$ . From Eq. (25), the  $\mathbf{D}^{-1} = (\mathbf{C}_{H\mathbf{x},H\mathbf{x}} + \mathbf{I})^{-1}$  matrix inverts  $\lambda_i + 1$  so that the largest eigenvalues become the smallest and vice versa. This means when  $\lambda_i$  is large, the contribution from  $\mathbf{C}_{H\mathbf{x},H\mathbf{x}}$  dominates (and therefore  $\frac{1}{\lambda_i + 1} \rightarrow 0$ ), while when  $\lambda_i$  is small, the contribution from  $\mathbf{R}$  dominates (so that  $\frac{1}{\lambda_i + 1} \rightarrow 1$ ). Thus, the inverse is a weighting of the relative uncertainties between the observation ensemble and the observation error covariances. When  $i > r_{\text{true}}$ , the weighing erroneously asserts that the contribution from the ensemble-estimated observation covariance is zero; that is, the ensemble covariance is *completely certain* of this mapping direction, when in fact according to  $\mathbf{C}_{H\mathbf{x},H\mathbf{x}}$  this direction is actually a linear combination of the other observations.

The problem comes about because, according to the modeled  $\mathbf{C}_{H\mathbf{x},H\mathbf{x}}$ , the covariance of  $N_{\text{obs}}$  can be described with only  $r_{\text{true}}$  linearly independent vectors; that is, there is some linear dependence between the observations. To fix this, let us conceptually assimilate only  $r \leq r_{\text{true}}$  principal component observations. Consider

$$\tilde{H}\mathbf{x} = \mathbf{U}_r^T H\mathbf{x}, \quad (29)$$

where  $\mathbf{U}_r$  is a matrix with columns of the first  $r$  eigenvectors of  $\mathbf{C}_{H\mathbf{x},H\mathbf{x}}$  ( $\mathbf{v}_i, i = 1, \dots, r$ ). Thus,

$$\text{cov}(\tilde{H}\mathbf{x}, \tilde{H}\mathbf{x}) = \mathbf{U}_r^T \mathbf{C}_{H\mathbf{x},H\mathbf{x}} \mathbf{U}_r = \mathbf{S}_r, \quad (30)$$

where  $\mathbf{S}_r$  is a diagonal matrix with the first  $r$  eigenvalues of  $\mathbf{C}_{H\mathbf{x},H\mathbf{x}}$ . Furthermore, for the observations  $\tilde{\mathbf{y}} = \mathbf{U}_r^T \mathbf{y}$ ,

$$\text{cov}(\tilde{\mathbf{y}} - \tilde{H}(\bar{\mathbf{x}}_f)) = \mathbf{U}_r^T \text{cov}(\mathbf{y} - H(\bar{\mathbf{x}}_f)) \mathbf{U}_r = \mathbf{U}_r^T \mathbf{R} \mathbf{U}_r = \mathbf{I}_r, \quad (31)$$

as  $\mathbf{R} = \mathbf{I}$  by the preprocessing step of Eq. (4). Therefore, the  $\mathbf{M}'$  (the updated  $\mathbf{M}$  with these  $\tilde{\mathbf{y}}$ ) from Eq. (5) is

$$\mathbf{M}' = (\mathbf{S}_r + \mathbf{I}_r + (\mathbf{S}_r + \mathbf{I}_r)^{1/2}), \quad (32)$$

a diagonal matrix, and so likewise is

$$\mathbf{D}' = (\mathbf{S}_r + \mathbf{I}_r). \quad (33)$$

The inverses of these matrices are now easy to compute as

$$(\mathbf{M}')^{-1} \mathbf{z}'_j = (\mathbf{S}_r + \mathbf{I}_r + (\mathbf{S}_r + \mathbf{I}_r)^{1/2})^{-1} \mathbf{U}_r^T \mathbf{z}_j. \quad (34)$$

Finally, while it is theoretically possible to directly work with  $\tilde{\mathbf{y}}$ , for observation-space covariance localization in  $\mathbf{C}_{H\mathbf{x},H\mathbf{x}}$ , we must project back to the observation space so that each observation has an ascribable location. Therefore, writing in terms of the summation notation in Eqs. (22) and (25) above,

$$\mathbf{U}_r (\mathbf{M}')^{-1} \mathbf{z}'_j = \sum_{i=1}^r \lambda'_i \alpha_{i,j} \mathbf{v}_i \quad (35)$$

and

$$\mathbf{U}_r (\mathbf{D}')^{-1} \mathbf{y}' = \sum_{i=1}^r \frac{\beta_i}{\lambda_i + 1} \mathbf{v}_i. \quad (36)$$

Note that the only differences between Eqs. (22) and (35) and (25) and (36) are the projection (noise) terms. Assimilating principal component observations removes these terms, and they would be zero if  $\mathbf{C}_{H\mathbf{x},H\mathbf{x}}$  was full rank. Therefore, assimilating principal component observations is equivalent to neglecting the projection terms in Eqs. (22) and (25), and thus we need not actually directly assimilate principal component observations; the inverse matrix action can be transformed round-trip back into observation space, avoiding the null space of  $\mathbf{C}_{H\mathbf{x},H\mathbf{x}}$  in the process. The projection-neglected Eqs. (35) and (36) are thus used with Eq. (26) to compute  $\tilde{\mathbf{K}}$  and  $\mathbf{K}$  in Eq. (1) to solve the ESRF problem.

Lastly, it is clear that with the eigenpair approach,  $r$  should be as close to  $r_{\text{true}}$  as possible, for the smallest eigenvalues become the most important, most certain directions to project along. This can pose numerical difficulties that our eigenproblem solver must manage. We will address this issue in section 3.

### c. Model-space localization

Equation (10) localizes covariance in observation-space, meaning observations are assumed to have an ascribable location, and distances are computed based

on the distances between these observations. However, unlike traditional observations such as dropsondes, integrated observations such as satellite radiances (which are affected by the entire vertical structure of the atmosphere and have a horizontal footprint based on the antenna size) do not have a single location. [Campbell et al. \(2010\)](#) demonstrated how the use of observation-space localization can degrade the quality of the analysis in these cases, while [Lei and Whitaker \(2015\)](#) showed that observation-space localization produces a superior analysis for a particular satellite radiance case. Both approaches are idealized models of how covariance decays (isotropically) as a function of distance, and which covariance model is superior is likely to be case dependent. This is an area of research that requires further investigation.

As we directly find the eigenpairs of the sparse matrix  $\mathbf{C}_{H\mathbf{x},H\mathbf{x}}$ , our method is theoretically compatible with model-space localization. To use this approach, the observation operator tangent-linear  $\mathbf{H}$  and adjoint  $\mathbf{H}^T$  are applied to the localized model covariance as

$$\mathbf{C}_{H\mathbf{x},H\mathbf{x}}^{\text{model}} = \mathbf{H}(\rho_{\mathbf{x},\mathbf{x}} \circ \mathbf{Q}_{\mathbf{x},\mathbf{x}})\mathbf{H}^T, \quad (37)$$

where

$$\mathbf{Q}_{\mathbf{x},\mathbf{x}} = \frac{\mathbf{X}_f'(\mathbf{X}_f')^T}{N_{\text{ens}} - 1}, \quad (38)$$

and

$$(\rho_{\mathbf{x},\mathbf{x}})_{i,j} = \ell(d_{i,j}|L_{i,j}), \quad (39)$$

where  $d_{i,j}$  is the distance between the location of two model states  $i$  and  $j$  with the same localization function as Eq. (11). Equation (26) is changed analogously as

$$\mathbf{C}_{\mathbf{x},H\mathbf{x}}^{\text{model}} = (\rho_{\mathbf{x},\mathbf{x}} \circ \mathbf{Q}_{\mathbf{x},\mathbf{x}})\mathbf{H}^T. \quad (40)$$

It is not necessary to form the prohibitively expensive full  $N_{\text{state}} \times N_{\text{state}}$   $\mathbf{Q}_{\mathbf{x},\mathbf{x}}$  matrix, as only the locations of  $\mathbf{H}$  and  $\mathbf{H}^T$  need to be computed in Eq. (38). This can be accomplished pairwise for each observation  $i$  and  $j$  by computing the localized model points at the observation location overlaps. However, in addition to requiring observation operator tangent linear models and adjoints, this approach requires some bookkeeping and communication to track which observation operators use which model points, and for ease of implementation in this paper, we therefore demonstrate our methodology with observation-space

localization without the loss of generality for the theoretical algorithm.

#### d. Comparison to other selected eigenvalue-based EnKF methods

Previous works have used eigenproblem-based methods to solve the EnKF equations based on similar principles but without, to the authors' best knowledge, a full accounting of covariance localization and rank deficiency. [Bishop et al. \(2001\)](#) uses the eigenvalue decomposition of the  $N_{\text{ens}} \times N_{\text{ens}}$  matrix  $\frac{1}{N_{\text{ens}}}(\mathbf{H}\mathbf{X})^T\mathbf{H}\mathbf{X}$  to find the solution of the inverse in Eq. (1). This is equivalent to finding the Sherman–Woodbury update as shown in [Tippett et al. \(2003\)](#). However, this approach cannot be used in the presence of covariance localization, as Eq. (10) includes the Schur product and increases the rank of the matrix  $\mathbf{C}_{H\mathbf{x},H\mathbf{x}}$ , requiring either computation of all nonsingular eigenvalues of  $\mathbf{C}_{H\mathbf{x},H\mathbf{x}}$  or the columns of the square root of the localization matrix as in [Bishop and Hodyss \(2009\)](#) using the so-called modulation product. [Posselt and Bishop \(2012\)](#) showed how to extend the findings of [Bishop et al. \(2001\)](#) in a computationally efficient way to where there are more ensemble members than observations making the  $\mathbf{C}_{H\mathbf{x},H\mathbf{x}}$  matrix full rank; we extend this analysis without this assumption. [Hamill and Snyder \(2002\)](#) used eigenvectors of  $\mathbf{C}_{H\mathbf{x},H\mathbf{x}}$  to test the impact of a single radiosonde on the mean forecast, assuming the matrix is full rank. Our paper likewise extends this to multiple (possibly rank deficient) observations with covariance localization, updating both the mean and ensemble. [Furrer and Bengtsson \(2007\)](#) examined the issue of covariance localization in the ensemble Kalman filter and found several approaches to estimate “taper matrices,” which approximate the localized covariance matrix, yet are not necessarily positive semidefinite. In this work we will directly consider the eigenpairs of the localized covariance matrix instead.

### 3. Direct solution of square root observation filter equations: Numerical implementation

In the previous section we developed a theoretical method for directly solving the ESRF Eq. (1) by computing the eigenpairs of  $\mathbf{C}_{H\mathbf{x},H\mathbf{x}}$  such that even a rank-deficient matrix can be utilized. Finding eigenpairs of a matrix is a prevalent problem across many different research fields (e.g., [Borsanyi et al. 2015](#); [Keçeli et al. 2016](#)) as well as industry (e.g., [Bryan and Leise 2006](#)). Therefore, we are able to benefit from the massive amounts of effort that has been expended in this area by choosing an appropriate library.

### Finding the eigenpairs of $\mathbf{C}_{H\mathbf{x},H\mathbf{x}}$

In the previous section we showed that finding the eigenvectors and eigenvalues of  $\mathbf{C}_{H\mathbf{x},H\mathbf{x}}$  was sufficient to solve the most difficult part of Eq. (1)—the matrix inverse of  $\mathbf{M}$ , which involves a square root of  $\mathbf{D}$ , which in turn involves a Schur product.

$\mathbf{C}_{H\mathbf{x},H\mathbf{x}}$  is a sparse  $N_{\text{obs}} \times N_{\text{obs}}$  matrix. In HEDAS, where frequent cycling is used, the number of assimilated observations in a single cycle is on the order of  $10^5$ , and therefore these matrices can be represented directly in memory spread across multiple machines. Even a full  $10^5 \times 10^5$  double-precision matrix would correspond approximately to 150 GB of memory, and with 32 GB of RAM per node, the matrix can be represented in memory distributed across approximately five or six computational nodes. Furthermore, because of covariance localization, these matrices are sparse enough that the covariance matrix can fit in memory on just one or at most two nodes. However, as the number of observations increases, it may become necessary to represent  $\mathbf{C}_{H\mathbf{x},H\mathbf{x}}$  only in terms of the matrix action upon some vector  $\mathbf{w}$  as  $\mathbf{C}_{H\mathbf{x},H\mathbf{x}}\mathbf{w}$ . Furthermore, as the data assimilation problem continues to grow, we would benefit enormously from the experience of the computational science community with different methods to store and solve sparse linear algebra and eigenvalue problems.

We wish to use a flexible solver that is capable of handling multiple types of problems (e.g., if we required using a matrix action at some point in the future) and to take advantage of different linear algebra techniques (such as direct and iterative matrix solvers). The Scalable Library for Eigenvalue Problem Computations (SLEPc; Hernández et al. 2003, 2005; Campos and Roman 2012; Roman et al. 2016), which is built upon the Portable, Extensible Toolkit for Scientific Computation (PETSc; Balay et al. 1997, 2016a,b), meets these requirements. PETSc recently celebrated its 20th anniversary (<http://www.mcs.anl.gov/petsc/petsc-20.html>), and SLEPc/PETSc has been used successfully by Keçeli et al. (2016) on over 266 000 cores to extract 330 000 eigenpairs in approximately 190 s. In short, it is a mature and flexible and scalable library with developmental support and a wide-ranging user base.

We now describe our parallel numerical implementation built upon this framework. We first must divide the computational domain among processing elements (PEs). We wish to allow all observation operators to be computed in parallel, so each PE must have all the model points necessary to compute the operator. With this in mind, while additional observations remain, the least-utilized PE is chosen, and the observation and all model points necessary to compute it are assigned to that PE (duplicating model indexes if necessary;

however, only one PE is assigned as the owner of a given model point). Once all observations have been distributed, the remaining model indexes not directly used in the computation of observation operators are divided among the remaining PEs. An attempt is made to keep vertical columns of the domain together on one PE to improve performance. For any given model point, then, multiple PEs may load the data but only one will be marked as the owner; for each observation, only one PE will be responsible for computing and transferring that observation.

For a given PE, the selected indices across all ensemble members are loaded from disk using Parallel netCDF (Li et al. 2003), and the forward observation calculations and quality control are performed in parallel and then broadcasted to all PEs. We now divide  $\mathbf{C}_{H\mathbf{x},H\mathbf{x}}$  into groups by assigning a starting and ending row for which a given PE will be responsible.

For our eigenvalue problem, we use the multiple shift-invert Lanczos method (Bai et al. 2000; Aktulga et al. 2014; Keçeli et al. 2016). A two-step process is used. The first stage divides the eigenvalue spectrum, in parallel, into regions with approximately equal numbers of eigenpairs. The second step computes all eigenvalues and eigenvectors, in parallel, within each region from the previous step. Each region is assigned multiple processors, allowing for two levels of parallelism.

Let  $n_{\text{regions}}$  be the number of regions and  $n_{\text{ppr}}$  be the number of PEs per region. Therefore,

$$n_{\text{PE}} = n_{\text{regions}} \times n_{\text{ppr}}, \quad (41)$$

where  $n_{\text{PE}}$  is the total number of PEs.

To compute the matrix spectrum of a matrix  $\mathbf{C}_{H\mathbf{x},H\mathbf{x}}$ , the matrix inertia theorem is utilized, which states that the number of eigenvalues in the region  $(\zeta, \eta)$ , for  $\zeta, \eta \in \mathbb{R}$ , can be found by decomposing [in our case, in a parallel and distributed way by calling from PETSc the Multifrontal Massively Parallel Sparse Direct Solver (MUMPS); Amestoy et al. 2000] both

$$\mathbf{C}_{H\mathbf{x},H\mathbf{x}} - \zeta \mathbf{I} = \mathbf{L}_{\zeta} \Delta_{\zeta} \mathbf{L}_{\zeta}^T \quad (42)$$

and

$$\mathbf{C}_{H\mathbf{x},H\mathbf{x}} - \eta \mathbf{I} = \mathbf{L}_{\eta} \Delta_{\eta} \mathbf{L}_{\eta}^T, \quad (43)$$

where  $\mathbf{L}_{\zeta}$  and  $\mathbf{L}_{\eta}$  are lower triangular and  $\Delta_{\zeta}$  and  $\Delta_{\eta}$  are block diagonal. The difference between negative entries of  $\Delta_{\zeta}$  and  $\Delta_{\eta}$  in Eqs. (42) and (43) is the number of eigenvalues in the region  $(\zeta, \eta)$ . As in Aktulga et al. (2014), an evenly spaced first guess of region divisions is given to each of the  $n_{\text{region}}$  groups, and the matrix inertia is computed at each of these points. The resulting



spectrum is then interpolated so that each region has an approximately equal number of eigenpairs for efficient division of effort. The matrix inertia at these divisions is computed once again to ensure the eigenvalue spectrum is exact. As mentioned in [Aktulga et al. \(2014\)](#), this spectrum slicing strategy is a computationally expensive step ultimately worth the cost in order to achieve efficient distribution of effort for what follows.

Once each processing region (with  $n_{\text{ppr}}$  PEs) is assigned a portion of the eigenvalue spectrum, the shift-invert Lanczos method is used within SLEPc to compute all of the known number of eigenvalues and eigenvectors within this region ([Campos and Roman 2012](#)). The left and right endpoints compute the smallest and largest eigenpairs, respectively, both of which SLEPc can handle. Because it is easier to work with a non-singular matrix, we compute the eigenpairs of  $\mathbf{C}_{H\mathbf{x},H\mathbf{x}} + \mathbf{I}$  and subtract 1 from the eigenvalues to give  $\lambda_i$ .

After the eigenpairs have been computed, Eqs. (35) and (36) are found and then distributed to each PE. Since  $\mathbf{A} = \mathbf{U}_r(\mathbf{M}')^{-1}\mathbf{Z}$  (where  $\mathbf{Z} = -H\mathbf{X}$ ) is  $(N_{\text{obs}} \times N_{\text{ens}})$  and  $\mathbf{b} = \mathbf{U}_r(\mathbf{D}')^{-1}\mathbf{Y}$  is  $(N_{\text{obs}} \times 1)$  [which again are computed as sums in the right-hand sides of (35) and (36)], these are combined into  $(\mathbf{A}|\mathbf{b})$ , a  $(N_{\text{obs}} \times N_{\text{ens}} + 1)$  matrix, which is relatively small. Each processing element then computes  $\mathbf{C}_{\mathbf{x},H\mathbf{x}}(\mathbf{A}|\mathbf{b})$  for each model index it is responsible for to solve Eq. (1) in an embarrassingly parallel way. This is possible, as each PE has the entire  $(\mathbf{A}|\mathbf{b})$  matrix and observation locations in memory, and only the local model analysis grid points are required to compute the localization and matrix product terms in Eq. (26) in  $\mathbf{C}_{\mathbf{x},H\mathbf{x}}(\mathbf{A}|\mathbf{b})$  with observation-space localization.<sup>4</sup> This is sufficient for each PE to solve the Kalman equations; that is, there is no need for any explicit communication among processors after  $(\mathbf{A}|\mathbf{b})$  has been distributed except to write the computed analysis to disk.

Our parallel algorithm for directly solving Eq. (1) is summarized in [Table 1](#).

#### 4. Numerical results

We implemented the abovementioned algorithm, including installing and testing the SLEPc/PETSc library, in Fortran on NOAA's Research and Development High Performance Computing Program Jet supercomputing system. We tested this with a Hurricane Edouard (2014)

TABLE 1. Parallel direct square root EnKF implementation pseudocode.

---

```

PROGRAM Parallel Direct  $\sqrt{\text{ENKF}}$  Solver
  DISTRIBUTE obs, model indexes for  $\bar{\mathbf{x}}_f, \mathbf{X}_f$ 
  COMPUTE quality control, fwd obs
  DIVIDE sparse  $\mathbf{C}_{H\mathbf{x},H\mathbf{x}}$  into  $n_{\text{regions}}, n_{\text{ppr}}$ 
  FIND eigenvalue spectrum using matrix
    inertia
  INTERPOLATE spectrum to distribute
    eigenpairs
  RECOMPUTE exact eigenvalue spectrum
  CALL eigensolver for all eigenpairs in
    region
  SOLVE  $\alpha_{ij}, \mathbf{A} = \mathbf{M}^{-1}\mathbf{Z}, \beta_i, \mathbf{b} = \mathbf{D}^{-1}\mathbf{Y}$ 
  DISTRIBUTE  $(\mathbf{A}|\mathbf{b})$  ( $N_{\text{obs}} \times N_{\text{ens}} + 1$ ) to CPUs
  FOR EACH model point  $i$  on each CPU, in ||
    COMPUTE  $(\bar{\mathbf{x}}_a)_i, (\mathbf{X}'_a)_i$ 
  WRITE  $\bar{\mathbf{x}}_a, \mathbf{X}'_a$  to disk
END Parallel Direct  $\sqrt{\text{ENKF}}$  Solver

```

---

case with 30 ensemble members and approximately 8-K quality-controlled observations from sources including satellite retrievals and the NASA AV-6 Global Hawk 20140916GH Storm Survey mission ([Zawislak et al. 2016](#); [Rogers et al. 2016](#)). The details of the evolution of the tropical cyclone can be found in [Stewart \(2014\)](#). As with other HEDAS studies, the initial conditions were initialized from the GFS ensemble 6h before assimilation; for more information see [section 1b](#).

We test assimilating these observations with a fixed covariance localization length  $L$  from Eqs. (27) and (11) set to  $L = 30, 60, 90, 120$ , and  $240$  as  $c = L/2$  from Eq. (4.10) of [Gaspari and Cohn \(1999\)](#). The unit on  $L$  is the number of grid cells, each of which is approximately 2 km, on the total domain of size  $232 \times 451 \times 60$  longitude/latitude/vertical model grid points, and  $L$  represents the number of horizontal grid points before the correlation reaches 0. Vertically the correlation length scale is  $L_v = 15$ ; in other words, the vertical distance is scaled by  $L/L_v$  so that after  $\pm 15$  vertical levels, the correlation goes to 0. While we vary the horizontal  $L$ ,  $L_v$  is fixed at 15.

The eigenvalue spectrum of these different cases is shown in [Fig. 1](#), which shows that the number of eigenvalues (i.e., the rank of  $\mathbf{C}_{H\mathbf{x},H\mathbf{x}}$ ) grows larger as  $L$  grows smaller. For all values of  $L$ , there is a sharp decay in the eigenvalue spectrum after  $10^{-2}$ ; likewise, there are few eigenvalues larger than  $10^2$ .

The prior forecast and analysis means based on Eqs. (35) and (36) are shown in [Fig. 2](#) for temperature in the midlevels (35 out of 60, where the first level is the surface and level 60 corresponds to 50 hPa). The prior demonstrates a symmetric warm core, while in this

<sup>4</sup>Note that with model-space localization, an additional processing step would be necessary to compute the intersection between the model grid points on each PE and the observation operators for  $\mathbf{C}_{\mathbf{x},H\mathbf{x}}$  in order to apply Eq. (40). This bookkeeping is the same as that used to compute  $\mathbf{C}_{H\mathbf{x},H\mathbf{x}}$  in Eq. (37).

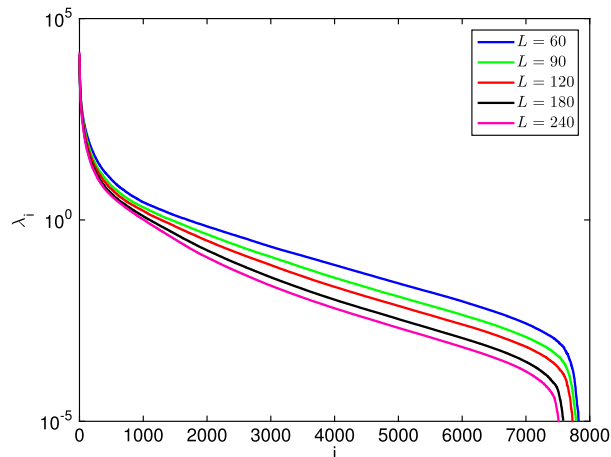


FIG. 1. Eigenvalue spectrum of  $\mathbf{C}_{Hx,Hx}$  as a function of  $L$  for the Hurricane Edouard assimilation test case. The number of eigenvalues (i.e., the rank of the matrix) grows as  $L$  shrinks, as expected.

case the analyzed vortex is not nearly as symmetrical. That the vortex may be unbalanced in the analysis is a common problem in data assimilation (Nehrkorn et al. 2015), and there are multiple potential remedies (Beezley and Mandel 2008; Kalnay and Yang 2010; Nehrkorn et al. 2015). The approach currently used in HEDAS is to distribute observations among frequently spaced assimilation subcycles using a storm-relative approach (Aksoy 2013), allowing the model dynamics to ease toward the observations over multiple assimilations. The results shown here are from the first and only such cycle, and therefore still include the deleterious effects of feature misalignment. However, as our intent is primarily to demonstrate the feasibility of this implementation on a relatively large set of observations, we do not account for these issues here.

Figure 3 shows the same analysis with the *rank-deficiency projection error* of the potentially rank-deficient observation covariance matrix. We define this projection error to be the effects of using Eqs. (22) and (25), which amounts to possibly erroneously assuming  $\mathbf{C}_{Hx,Hx}$  is full rank, instead of the corresponding corrected versions Eqs. (35) and (36), respectively, that address this rank deficiency. The rank-deficiency projection error grows larger as a function of  $L$  (the correlation length scale) as the rank of the matrix decreases and the impact of each observation is over a larger part of the domain. Figure 4 shows the  $\log_{10}$  absolute difference between the rank-deficiency-neglected analysis and the corrected analysis of the 10-m wind speed, first arising from differences in ordering (the difference is zero to double-precision machine  $\epsilon$ ), and then as a function of correlation length scale. The features of this projection error also suggest how the eigenpairs (in this

case the extraneous linearly dependent directions) exhibit complex spatial structure.

Figure 5 shows the root-mean-square rank-deficiency projection error (at all points over the entire domain) and maximum projection error as a function of model level (the first level is the surface) for temperature, water vapor, and condensed water mass for our Edouard case for the different  $L$  values. For the longest  $L$ , the RMSE can reach  $0.55 \text{ K}$ ,  $0.65 \text{ g kg}^{-1}$ , and  $0.25 \text{ g kg}^{-1}$  for temperature, water vapor, and condensed water mass, respectively, in the middle to upper troposphere. The maximum projection error arising from the projection onto degenerate directions can reach values on the order of approximately 5 times the RMSE. In short, the projection error can be surprisingly large and likely physically meaningful, especially with larger values of  $L$ .

The eigenvalue computations took between 25 s ( $L = 30$ ) and 120 s ( $L = 240$ ) with 384 PEs—in other words, fewer eigenvalues actually take longer to compute. This is because the matrix becomes more dense as  $L$  grows and working with the matrix takes longer. In all the cases tested, the error of  $\|\mathbf{C}_{Hx,Hx}\mathbf{v}_i - \lambda_i\mathbf{v}_i\|$  was less than  $10^{-7}$  for all  $i$ .

We conducted our tests on the NOAA Jet supercomputing system (tjet processing queue) where each node has 12 cores with a 2.66-GHz CPU and 2 GB of RAM, each connected via a quad data rate (QDR) InfiniBand. Tests determined that  $n_{\text{ppr}} = 6$  (six PEs per region) performed best. The peak flops/node is 127.7 gigaflops, while in the SLEPc portion of the code we were able to achieve a peak performance of approximately 440 gigaflops with 384 PEs with this setup—about 11% of the theoretically optimal performance. Figure 6 shows the increase in computation speed as a function of the number of PEs with a fixed  $L = 60$ . As shown, the time to distribute the observations and model indices across the PEs (Fig. 6a) is a fixed time, as it is currently done in serial. Parallel netCDF read performance (Fig. 6b) showed improvement as the number of PEs increased, but parallel writes (Fig. 6c) showed a marked decrease in performance as the PEs competed with one another for access to the disk head. The eigenproblem computation (Fig. 6c) and filtering time (Fig. 6d) [i.e., solution of Eq. (1) once  $(\mathbf{A}|\mathbf{b})$  has been distributed to all PEs] increase even faster than linearly as a result of the advantages of the smaller matrix sizes fitting in cache memory. Therefore, overall (Fig. 6f) the problem is input/output (I/O) bound, and improvements to the parallel file system would be necessary for additional performance gains. This makes sense, as the problem of reading and writing all  $N_{\text{state}} \times N_{\text{ens}}$  variables is by far the most expensive operation in data assimilation. An additional order of magnitude of observations than our current test would likely necessitate additional PEs, but for the

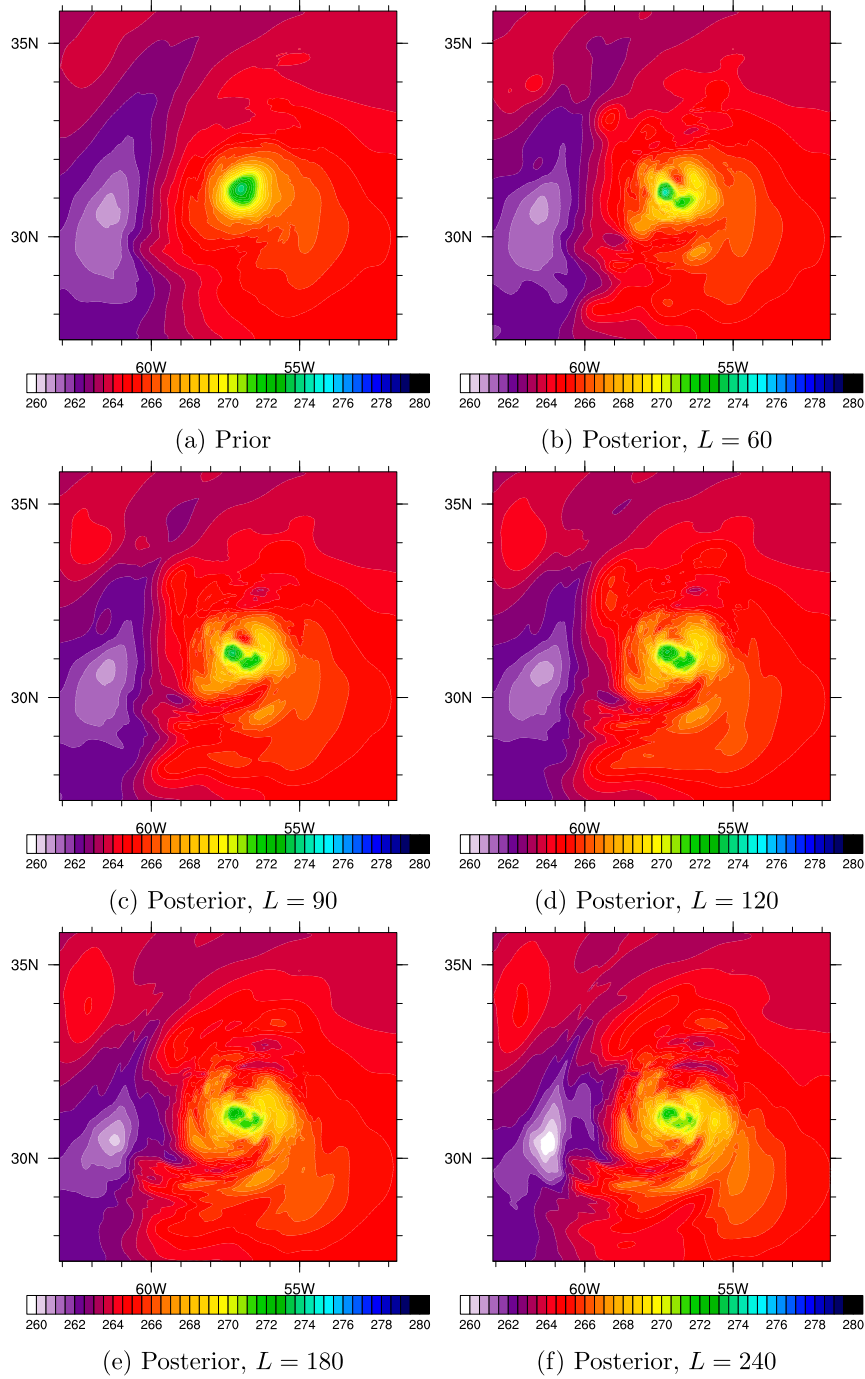


FIG. 2. Temperature at level 35 (of 60 total, corresponding to a height of approximate 5.5 km) for (a) the prior mean, (b)–(f) analysis means with  $L = 60$ ,  $L = 90$ ,  $L = 120$ ,  $L = 180$ , and  $L = 240$ . Analysis uses Eqs. (35) and (36) to remove the projection error arising from neglecting the rank deficiency of  $\mathbf{C}_{H\mathbf{x}, H\mathbf{x}}$ .

current problem size ( $\approx 10^4$ ), 384 PEs achieve the maximum parallel performance gain.

From these results and other results with the SLEPc library (e.g., Keçeli et al. 2016), we have every reason

to believe this parallel direct method will be able to scale  $10^6 - 10^7$  observations—or more given additional computational resources, especially with further improvements to the parallel file system.

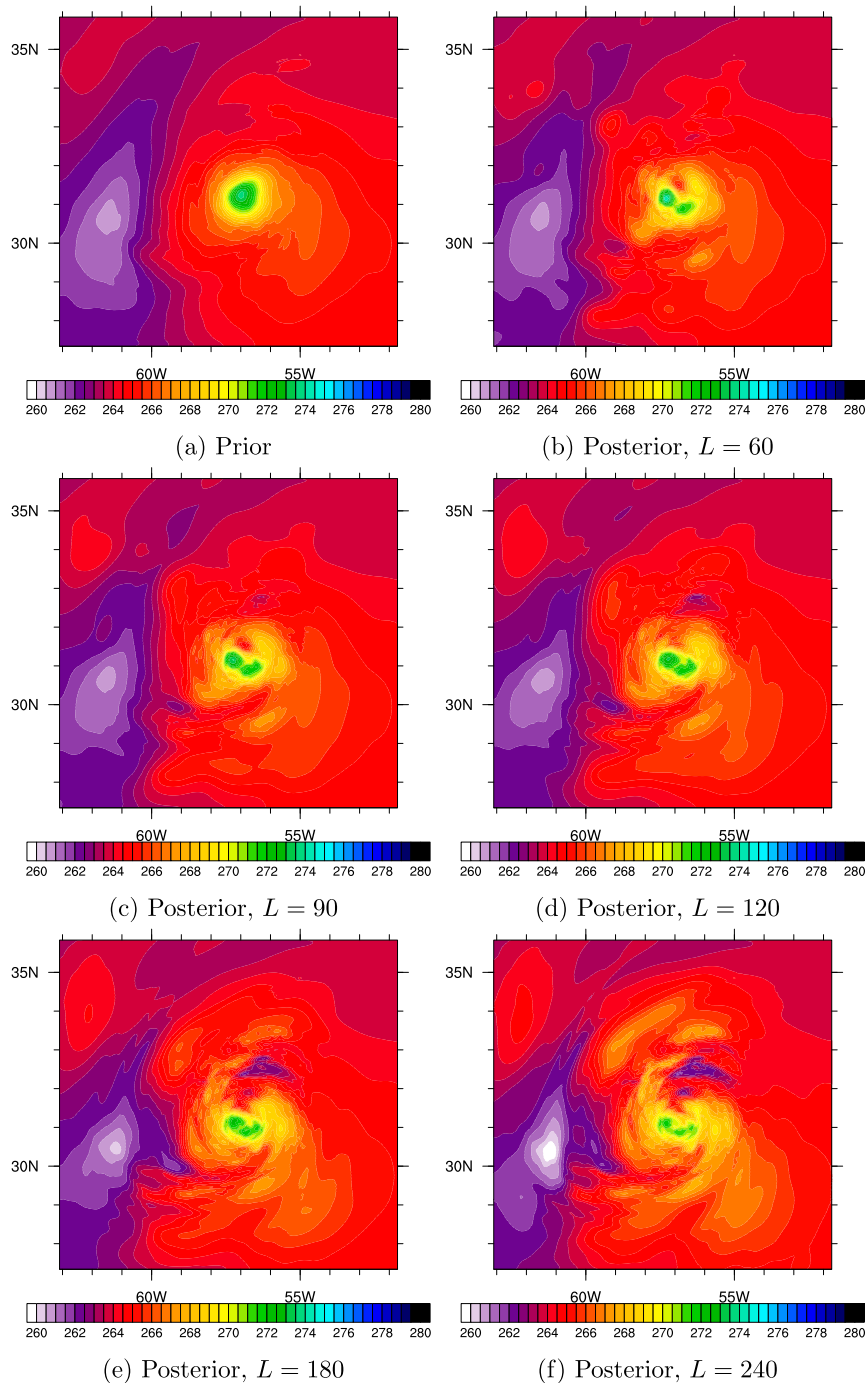


FIG. 3. As in Fig. 2, but without accounting for the rank-deficiency projection error from Eqs. (22) and (25). Analysis is nearly identical for  $L = 60$  and  $L = 90$ , but by  $L = 240$  there are large differences. Rank-deficiency projection error grows with  $L$  and the number of observations.

## 5. Conclusions

In this paper we describe an algorithm to compute, in parallel, a direct solution to the ensemble square root Kalman filter (ESRF) equations in a way that does not depend on the ordering of observations. The previous

approach for solving the ESRF equations, based on a serial assimilation of observations, became suboptimal when covariance localization was introduced, as the covariance localization operation is not linear and does not commute with serial assimilation in ESRF implementations described to date. While a consistent

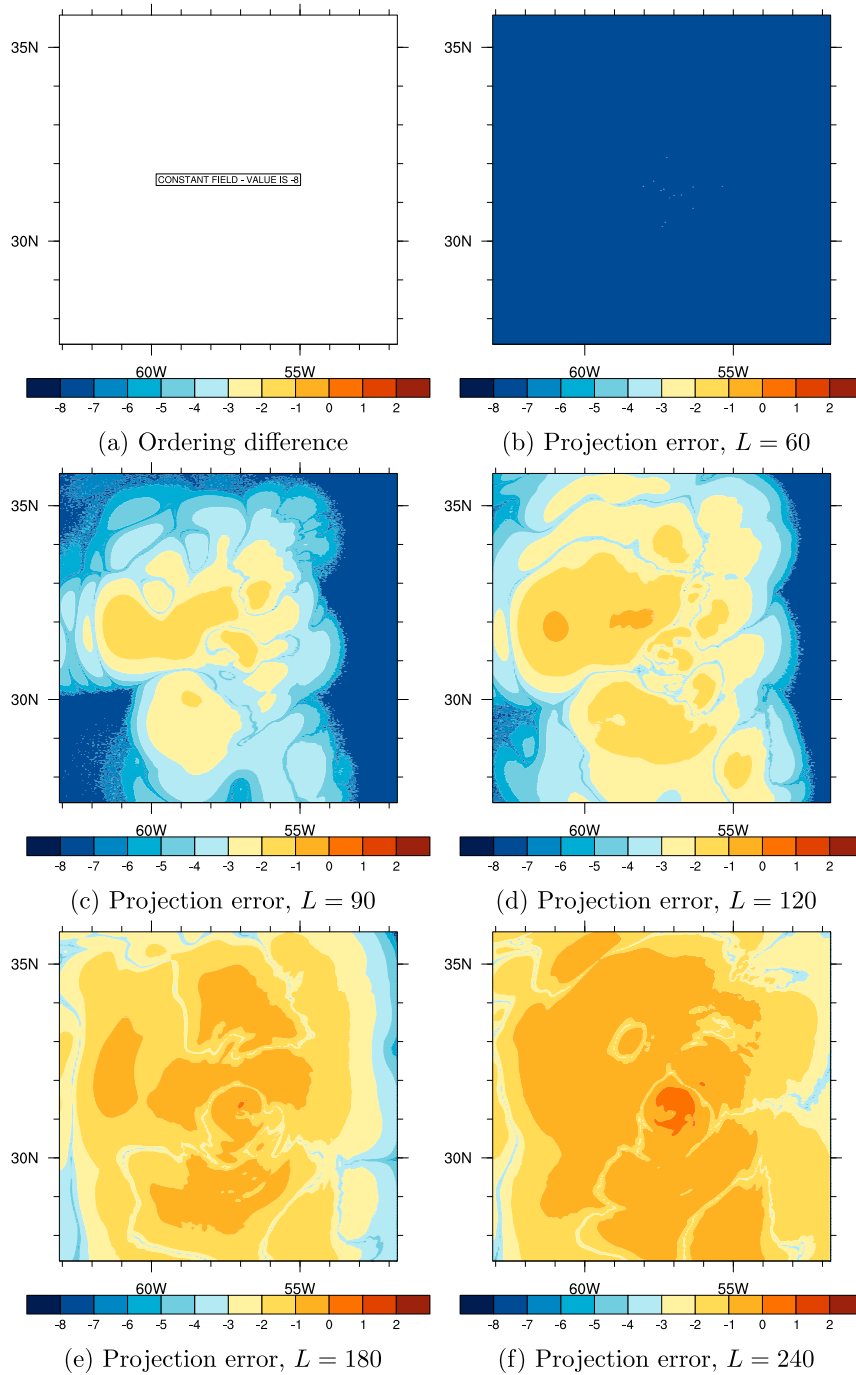


FIG. 4. The log<sub>10</sub> absolute difference between analyzed 10-m wind speed (m s<sup>-1</sup>). (a) Difference between two random orderings with all values of  $L$  is less than  $10^{-8}$ . (b)–(f) Difference between the simulations in Figs. 2 and 3. Rank-deficiency projection error grows as a function of  $L$  both in size and spread.

serial non-square root EnKF was previously presented in the CHEF implementation of Bishop et al. (2015) with the use of perturbed observations, this work describes an equivalently consistent parallel

direct ESRF implementation without the use of perturbed observations. To maintain consistency, we solve the ensemble square root Kalman filter equations directly in order to regain the optimal



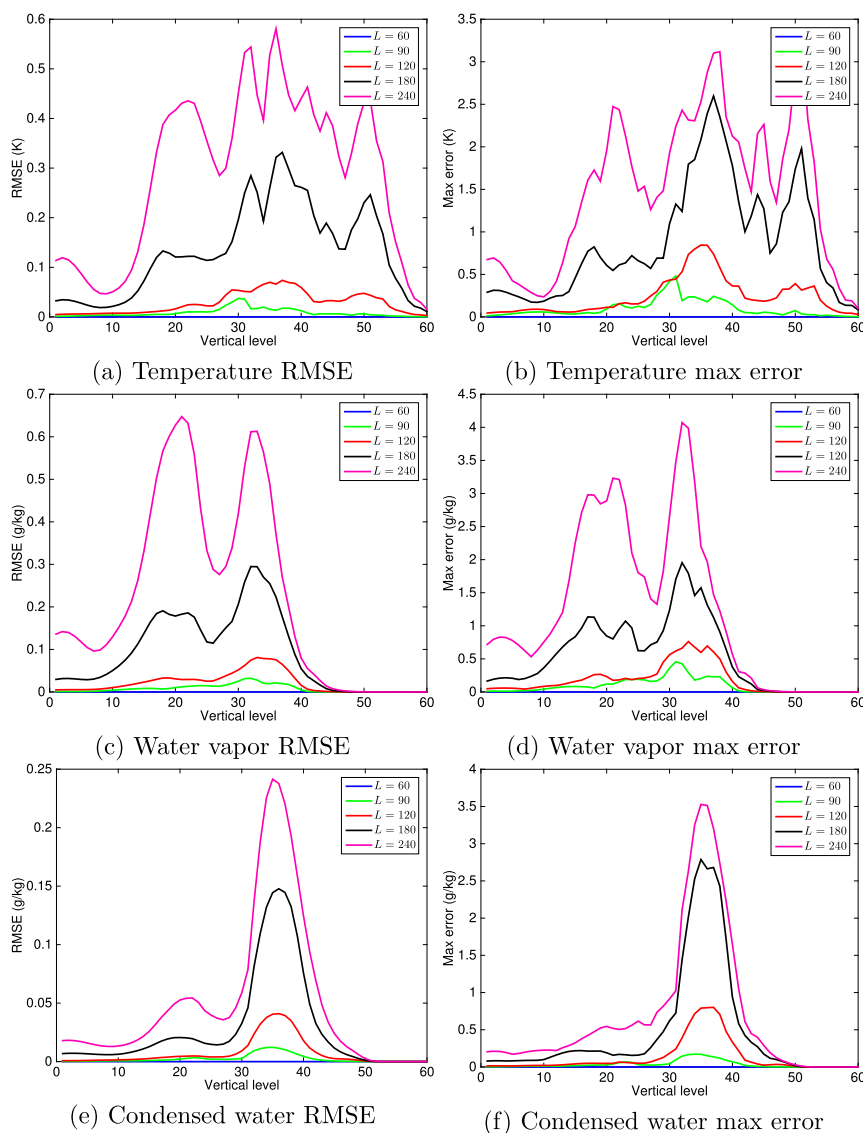


FIG. 5. Domain-averaged root-mean square projection error (RMSE) and maximum projection error (max error) for temperature, water vapor, and condensed water mass in the Hurricane Edouard case. RMSE for a given level is measured over the entire domain. Level 1 is the surface; 60 is the model top at 50 hPa. For this assimilation case, the error is largest near level 35 and reaches a maximum of 3 K, 4 g kg<sup>-1</sup>, and 3.5 g kg<sup>-1</sup> for these fields when  $L = 240$ . This rank-deficiency projection error is not present when using Eqs. (35) and (36).

properties of the filter. We demonstrate how this can be done in a practical manner by using the theory of matrix functions to show that for the observation covariance matrix  $\mathbf{C}_{Hx,Hx}$  (which can include covariance localization), after making the observation error covariance white by the transformation  $\mathbf{y} = \mathbf{R}_{old}^{1/2} \mathbf{y}_{old}$ , the eigenvectors  $\mathbf{v}_i$  of  $f(\mathbf{C}_{Hx,Hx})$  are the same as  $\mathbf{C}_{Hx,Hx}$  and the eigenvalues are the same as those found by “scalarizing” the function  $f$  and

substituting  $\lambda_i$ , that is,  $f(\lambda_i)$ . We demonstrate an error bound on the approximation. When using the first implementation described that explicitly considers the null space of the forward observation covariance, the solution contains a “projection error” when this matrix is rank deficient. We demonstrate how this projection error can be eliminated by conceptually assimilating principal component observations; in fact, this is exactly the same as removing the

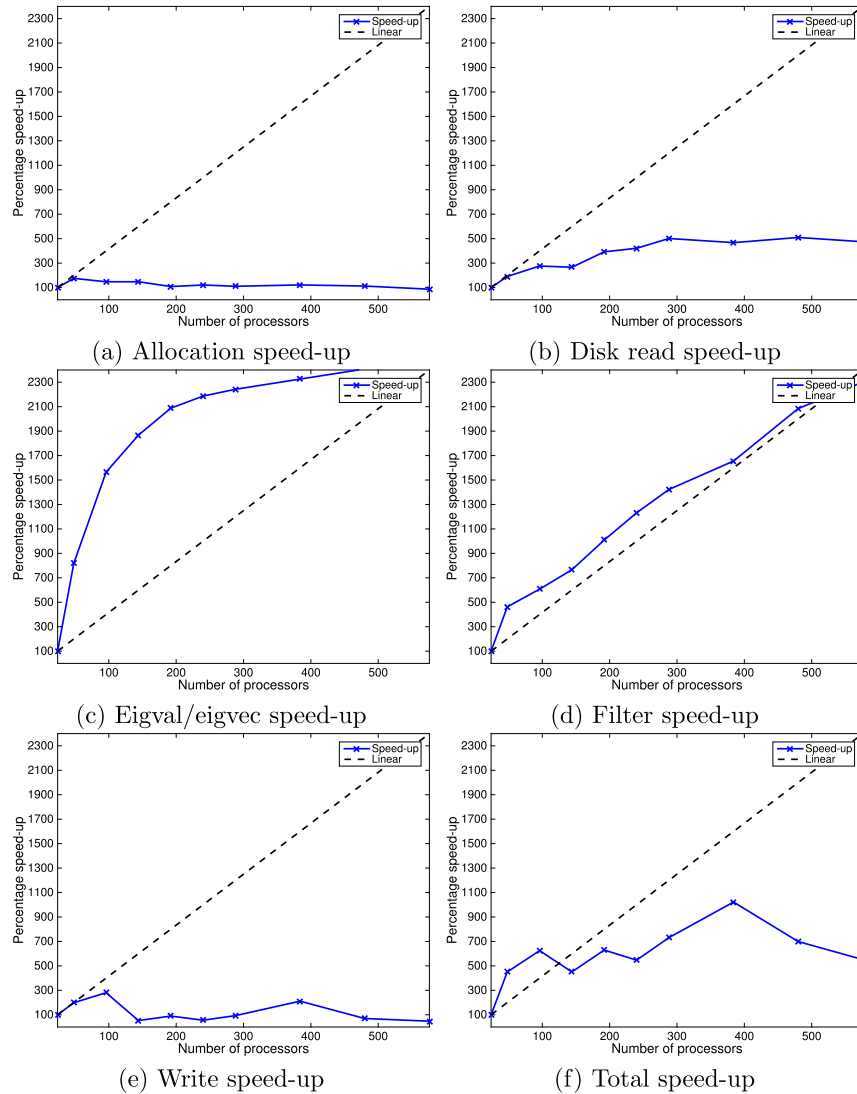


FIG. 6. Increase in speed of the parallel solution of the assimilation problem as a function of processing elements for  $L = 60$ . Allocation of indices is currently a fixed overhead, as it is done in serial. Parallel reads can increase in speed with additional PEs but only to approximately 4–5 times speedup. Eigenproblem computation is superlinear, as each region requires fewer eigenpairs to solve. The filtering step also grows slightly superlinearly as the matrices for each region fit into memory cache. Writing performance decreases with the number of PEs as a result of disk-head competition and shows large fluctuations apparently caused by system usage. Completion of 384 PEs takes approximately 8 min, which is approximately 10 times faster in total than 24 PEs. Overall, the process is I/O bound, so increasing beyond 384 PEs actually decreases the total performance.

projection term from the full solution. The rank-deficiency projection error increases with localization length but can be avoided entirely by the approach we describe. We demonstrate a successful parallel and scalable numerical implementation of this algorithm and demonstrate how the assimilation behaves as a function of covariance localization length.

As first noted in Bishop et al. (2001) for the ensemble transform Kalman filter (ETKF), for the actual assimilation method we require only the matrix product of the matrices considered here. This fact leads to two realizations. First, it may be feasible to construct a method that avoids computing all of the eigenpairs and instead uses matrix functions to solve the matrix products in a much faster way by considering only the space spanned

by the range of the most  $N_{\text{ens}} - 1$  matrix product directions rather than the entire eigenspace. This approach would be far superior to computing the full eigenspace. Second, the question of the null space of the forward observation covariance matrix, while theoretically interesting, may not be a practical issue for most EnKF implementations unless the right-hand product terms have a nonzero projection into the null space of this matrix. Additional research is necessary to determine whether and when this “rank projection” error occurs when using localized covariance matrices.

The approach described in this paper is more complex than the serial ESRF approximation, but once integrated with an appropriate library for solving a large, sparse, symmetric, positive semidefinite eigenproblem (taking advantage of the pervasive nature of the eigenproblem across disciplines), the approach is not conceptually difficult and becomes highly amenable to a scalable parallel implementation. These libraries can also integrate many different computational best practices and bring the solution of the ESRF equation into line with the state-of-the-art of numerical linear algebra. In this approach, because observations are assimilated all at once, unlike the approach given in Anderson and Collins (2007), the observations do not need to be treated as augmented state variables for the parallel implementation. Once the eigenproblem has been solved, when used with observation-space localization as presented here, the remaining analysis becomes embarrassingly parallel, as there is no need to communicate across processing elements. The parallel direct ensemble square root Kalman filter approach described here is therefore a new class of unperturbed observation direct ensemble square root Kalman filter that is amenable to large-scale parallel implementations.

**Acknowledgments.** The authors acknowledge the vital contributions of two anonymous reviewers. This research was partially funded by the NOAA Hurricane Forecast Improvement Project Award NA14NWS4680022. This research was carried out in part under the auspices of CIMAS, a joint institute of the University of Miami and NOAA (Cooperative Agreement NA67RJ0149). The authors acknowledge the NOAA Research and Development High Performance Computing Program for providing computing and storage resources that have contributed to the research results reported within this paper (<http://rdhpcs.noaa.gov>). We gratefully acknowledge the extensive technical support of Dr. José Roman.

## APPENDIX

TABLE A1. Symbols.

Symbol	Definition
$N_{\text{obs}}$	No. of observations
$N_{\text{ens}}$	No. of ensemble members
$\bar{\mathbf{x}}_f$	Forecast ensemble mean
$\bar{\mathbf{x}}_a$	Analysis ensemble mean
$\mathbf{K}$	Mean Kalman gain
$\mathbf{H}\mathbf{X}$	Forward observation perturbations
$\mathbf{C}_{\text{x,x}}$	Prior model covariance
$\mathbf{C}_{\text{Hx,Hx}}$	Forward observation covariance
$\mathbf{v}_i$	Eigenvector No. $i$ of $\mathbf{C}_{\text{Hx,Hx}}$
$\mathbf{D}$	$\mathbf{C}_{\text{Hx,Hx}} + \mathbf{R}$
$\lambda'_i$	Eigenvalue No. $i$ of $\mathbf{M}^{-1}$
$\alpha_{ij}$	$\mathbf{v}_i^T \mathbf{Z}_j$
$N_{\text{state}}$	No. of state variables
$\mathbf{y}$	Observations (normalized)
$\mathbf{X}'_f$	Forecast ensemble perturbations
$\mathbf{X}'_a$	Analysis ensemble perturbations
$\mathbf{K}$	Perturbations Kalman gain
$\mathbf{R} = \mathbf{I}$	Observation error covariance (normalized)
$\mathbf{C}_{\text{x,Hx}}$	Forward observation/model cross covariance
$\lambda_i$	Eigenvalue No. $i$ of $\mathbf{C}_{\text{Hx,Hx}}$
$\mathbf{Y}$	$\mathbf{y} - H(\bar{\mathbf{x}}_f)$
$\mathbf{M}$	$\mathbf{D} + \sqrt{\mathbf{D}}$
$\mathbf{Z}$	$\mathbf{Z} = -\mathbf{H}\mathbf{X}$ , $\mathbf{Z}_j$ is column $j$
$\beta_i$	$\mathbf{v}_i^T \mathbf{Y}$

## REFERENCES

- Aberson, S. D., A. Aksoy, K. J. Sellwood, T. Vukicevic, and X. Zhang, 2015: Assimilation of high-resolution tropical cyclone observations with an ensemble Kalman filter using HEDAS: Evaluation of 2008–11 HWRF forecasts. *Mon. Wea. Rev.*, **143**, 511–523, doi:10.1175/MWR-D-14-00138.1.
- Aksoy, A., 2013: Storm-relative observations in tropical cyclone data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.*, **141**, 506–522, doi:10.1175/MWR-D-12-00094.1.
- , S. Loruso, T. Vukicevic, K. J. Sellwood, S. D. Aberson, and F. Zhang, 2012: The HWRF Hurricane Ensemble Data Assimilation System (HEDAS) for high-resolution data: The impact of airborne Doppler radar observations in an OSSE. *Mon. Wea. Rev.*, **140**, 1843–1862, doi:10.1175/MWR-D-11-00212.1.
- , S. D. Aberson, T. Vukicevic, K. J. Sellwood, S. Loruso, and X. Zhang, 2013: Assimilation of high-resolution tropical cyclone observations with an ensemble Kalman filter using NOAA/AOML/HRD’s HEDAS: Evaluation of the 2008–11 vortex-scale analyses. *Mon. Wea. Rev.*, **141**, 1842–1865, doi:10.1175/MWR-D-12-00194.1.
- Aktulga, H. M., L. Lin, C. Haine, E. G. Ng, and C. Yang, 2014: Parallel eigenvalue calculation based on multiple shift-invert Lanczos and contour integral based spectral projection method. *Parallel Comput.*, **40**, 195–212, doi:10.1016/j.parco.2014.03.002.
- Amestoy, P. R., I. S. Duff, and J. Y. L’Excellent, 2000: Multifrontal parallel distributed symmetric and unsymmetric solvers. *Comput. Methods Appl. Mech. Eng.*, **184**, 501–520, doi:10.1016/S0045-7825(99)00242-X.

- Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903, doi:[10.1175/1520-0493\(2001\)129<2884:AEAKFF>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2).
- , and N. Collins, 2007: Scalable implementations of ensemble filter algorithms for data assimilation. *J. Atmos. Oceanic Technol.*, **24**, 1452–1463, doi:[10.1175/JTECH2049.1](https://doi.org/10.1175/JTECH2049.1).
- Bai, Z., J. Demmel, J. Dongarra, A. Ruhe, and H. van der Vorst, Eds., 2000: *Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide*, Software, Environments and Tools, Vol. 11, SIAM, 410 pp.
- Balay, S., W. D. Gropp, L. C. McInnes, and B. F. Smith, 1997: Efficient management of parallelism in object-oriented numerical software libraries. *Modern Software Tools for Scientific Computing*, Springer, 163–202.
- , and Coauthors, 2016a: PETSc users manual. Revision 3.7, Argonne National Laboratory Tech. Rep. ANL-95/11 Rev 3.7, 241 pp. [Available online at <http://www.mcs.anl.gov/petsc/petsc-current/docs/manual.pdf>.]
- , and Coauthors, 2016b: Portable, Extensible Toolkit for Scientific Computation. [Available online at <https://www.mcs.anl.gov/petsc/>.]
- Beezley, J. D., and J. Mandel, 2008: Morphing ensemble Kalman filters. *Tellus*, **60A**, 131–140, doi:[10.1111/j.1600-0870.2007.00275.x](https://doi.org/10.1111/j.1600-0870.2007.00275.x).
- Bishop, C. H., and D. Hodyss, 2009: Ensemble covariances adaptively localized with ECO-RAP. Part 2: A strategy for the atmosphere. *Tellus*, **61A**, 97–111, doi:[10.1111/j.1600-0870.2008.00372.x](https://doi.org/10.1111/j.1600-0870.2008.00372.x).
- , B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, **129**, 420–436, doi:[10.1175/1520-0493\(2001\)129<0420:ASWTET>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2).
- , B. Huang, and X. Wang, 2015: A nonvariational consistent hybrid ensemble filter. *Mon. Wea. Rev.*, **143**, 5073–5090, doi:[10.1175/MWR-D-14-00391.1](https://doi.org/10.1175/MWR-D-14-00391.1).
- Borsanyi, S., and Coauthors, 2015: Ab initio calculation of the neutron-proton mass difference. *Science*, **347**, 1452–1455, doi:[10.1126/science.1257050](https://doi.org/10.1126/science.1257050).
- Bryan, K., and T. Leise, 2006: The \$25,000,000,000 eigenvector: The linear algebra behind Google. *SIAM Rev.*, **48**, 569–581, doi:[10.1137/050623280](https://doi.org/10.1137/050623280).
- Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719–1724, doi:[10.1175/1520-0493\(1998\)126<1719:ASITEK>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2).
- Campbell, W. F., C. H. Bishop, and D. Hodyss, 2010: Vertical covariance localization for satellite radiances in ensemble Kalman filters. *Mon. Wea. Rev.*, **138**, 282–290, doi:[10.1175/2009MWR3017.1](https://doi.org/10.1175/2009MWR3017.1).
- Campos, C., and J. E. Roman, 2012: Strategies for spectrum slicing based on restarted Lanczos methods. *Numer. Algorithms*, **60**, 279–295, doi:[10.1007/s11075-012-9564-z](https://doi.org/10.1007/s11075-012-9564-z).
- Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10 143–10 162, doi:[10.1029/94JC00572](https://doi.org/10.1029/94JC00572).
- Furrer, R., and T. Bengtsson, 2007: Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. *J. Multivar. Anal.*, **98**, 227–255, doi:[10.1016/j.jmva.2006.08.003](https://doi.org/10.1016/j.jmva.2006.08.003).
- Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757, doi:[10.1002/qj.49712555417](https://doi.org/10.1002/qj.49712555417).
- Hamill, T. M., and C. Snyder, 2002: Using improved background-error covariances from an ensemble Kalman filter for adaptive observations. *Mon. Wea. Rev.*, **130**, 1552–1572, doi:[10.1175/1520-0493\(2002\)130<1552:UIBECF>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<1552:UIBECF>2.0.CO;2).
- , J. S. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.*, **129**, 2776–2790, doi:[10.1175/1520-0493\(2001\)129<2776:DDFOBE>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2).
- Hernández, V., J. E. Roman, and V. Vidal, 2003: SLEPc: Scalable Library for Eigenvalue Problem Computations. *High Performance Computing for Computational Science—VECPAR 2002*, Lecture Notes in Computer Science, Vol. 2565, Springer, 377–391, doi:[10.1007/3-540-36569-9\\_25](https://doi.org/10.1007/3-540-36569-9_25).
- , —, and —, 2005: SLEPc: A scalable and flexible toolkit for the solution of eigenvalue problems. *ACM Trans. Math. Software*, **31**, 351–362, doi:[10.1145/1089014.1089019](https://doi.org/10.1145/1089014.1089019).
- Higham, N. J., 2008: *Functions of Matrices: Theory and Computation*. Other Titles in Applied Mathematics, SIAM, 413 pp., doi:[10.1137/1.9780898717778](https://doi.org/10.1137/1.9780898717778).
- Holland, B., and X. Wang, 2013: Effects of sequential or simultaneous assimilation of observations and localization methods on the performance of the ensemble Kalman filter. *Quart. J. Roy. Meteor. Soc.*, **139**, 758–770, doi:[10.1002/qj.2006](https://doi.org/10.1002/qj.2006).
- Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811, doi:[10.1175/1520-0493\(1998\)126<0796:DAUAEK>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2).
- , and —, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **129**, 123–137, doi:[10.1175/1520-0493\(2001\)129<0123:ASEKFF>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2).
- Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *J. Basic Eng.*, **82**, 35–45, doi:[10.1115/1.3662552](https://doi.org/10.1115/1.3662552).
- Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 341 pp.
- , and S.-C. Yang, 2010: Accelerating the spin-up of Ensemble Kalman Filtering. *Quart. J. Roy. Meteor. Soc.*, **136**, 1644–1651, doi:[10.1002/qj.652](https://doi.org/10.1002/qj.652).
- Keçeli, M., H. Zhang, P. Zapol, D. A. Dixon, and A. F. Wagner, 2016: Shift-and-invert parallel spectral transformation eigensolver: Massively parallel performance for density-functional based tight-binding. *J. Comput. Chem.*, **37**, 448–459, doi:[10.1002/jcc.24254](https://doi.org/10.1002/jcc.24254).
- Keppenne, C. L., and M. M. Rienecker, 2002: Initial testing of a massively parallel ensemble Kalman filter with the Poseidon isopycnal ocean general circulation model. *Mon. Wea. Rev.*, **130**, 2951–2965, doi:[10.1175/1520-0493\(2002\)130<2951:ITOAMP>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<2951:ITOAMP>2.0.CO;2).
- Lahoz, W. A., and P. Schneider, 2014: Data assimilation: Making sense of Earth observation. *Front. Environ. Sci.*, **2**, doi:[10.3389/fenvs.2014.00016](https://doi.org/10.3389/fenvs.2014.00016).
- Lei, L., and J. S. Whitaker, 2015: Model space localization is not always better than observation space localization for assimilation of satellite radiances. *Mon. Wea. Rev.*, **143**, 3948–3955, doi:[10.1175/MWR-D-14-00413.1](https://doi.org/10.1175/MWR-D-14-00413.1).
- Li, J., and Coauthors, 2003: Parallel netCDF: A high-performance scientific I/O interface. *SC2003: Igniting Innovation; Proceedings of the ACM/IEEE SC2003 Conference*, 39, doi:[10.1109/SC.2003.10053](https://doi.org/10.1109/SC.2003.10053).
- Nehrkorn, T., B. K. Woods, R. N. Hoffman, and T. Aulign, 2015: Correcting for position errors in variational data assimilation. *Mon. Wea. Rev.*, **143**, 1368–1381, doi:[10.1175/MWR-D-14-00127.1](https://doi.org/10.1175/MWR-D-14-00127.1).
- Nerger, L., 2015: On serial observation processing in localized ensemble Kalman filters. *Mon. Wea. Rev.*, **143**, 1554–1567, doi:[10.1175/MWR-D-14-00182.1](https://doi.org/10.1175/MWR-D-14-00182.1).
- Posselt, D. J., and C. H. Bishop, 2012: Nonlinear parameter estimation: Comparison of an ensemble Kalman smoother with a Markov chain Monte Carlo algorithm. *Mon. Wea. Rev.*, **140**, 1957–1974, doi:[10.1175/MWR-D-11-00242.1](https://doi.org/10.1175/MWR-D-11-00242.1).

- Rogers, R. F., J. A. Zhang, J. Zawislak, H. Jiang, G. R. Alvey, E. J. Zipser, and S. N. Stevenson, 2016: Observations of the structure and evolution of Hurricane Edouard (2014) during intensity change. Part II: Kinematic structure and the distribution of deep convection. *Mon. Wea. Rev.*, **144**, 3355–3376, doi:[10.1175/MWR-D-16-0017.1](https://doi.org/10.1175/MWR-D-16-0017.1).
- Roman, J. E., C. Campos, E. Romero, and A. Tomás, 2016: SLEPc users manual: Scalable Library for Eigenvalue Problem Computations. Revision 3.7, Tech. Rep. DSIC-II/24/02, Departamento de Sistemas Informáticos y Computación, Universitat Politècnica de València, 116 pp.
- Stewart, S., 2014: National Hurricane Center tropical cyclone report: Hurricane Edouardo (AL062014) 11–19 September 2014. National Hurricane Center Rep., 19 pp. [Available online at [http://www.nhc.noaa.gov/data/tcr/AL062014\\_Edouard.pdf](http://www.nhc.noaa.gov/data/tcr/AL062014_Edouard.pdf).]
- Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square root filters. *Mon. Wea. Rev.*, **131**, 1485–1490, doi:[10.1175/1520-0493\(2003\)131<1485:ESRF>2.0.CO;2](https://doi.org/10.1175/1520-0493(2003)131<1485:ESRF>2.0.CO;2).
- Vukicevic, T., A. Aksoy, P. Reasor, S. D. Aberson, K. J. Sellwood, and F. Marks, 2013: Joint impact of forecast tendency and state error biases in ensemble Kalman filter data assimilation of inner-core tropical cyclone observations. *Mon. Wea. Rev.*, **141**, 2992–3006, doi:[10.1175/MWR-D-12-00211.1](https://doi.org/10.1175/MWR-D-12-00211.1).
- Wang, Y., Y. Jung, T. A. Supinie, and M. Xue, 2013: A hybrid MPI–OpenMP parallel algorithm and performance analysis for an ensemble square root filter designed for multiscale observations. *J. Atmos. Oceanic Technol.*, **30**, 1382–1397, doi:[10.1175/JTECH-D-12-00165.1](https://doi.org/10.1175/JTECH-D-12-00165.1).
- Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924, doi:[10.1175/1520-0493\(2002\)130<1913:EDAWPO>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2).
- Zawislak, J., H. Jiang, G. R. Alvey, E. J. Zipser, R. F. Rogers, J. A. Zhang, and S. N. Stevenson, 2016: Observations of the structure and evolution of Hurricane Edouard (2014) during intensity change. Part I: Relationship between the thermodynamic structure and precipitation. *Mon. Wea. Rev.*, **144**, 3333–3354, doi:[10.1175/MWR-D-16-0018.1](https://doi.org/10.1175/MWR-D-16-0018.1).
- Zhang, S., M. J. Harrison, A. T. Wittenberg, A. Rosati, J. L. Anderson, and V. Balaji, 2005: Initialization of an ENSO forecast system using a parallelized ensemble filter. *Mon. Wea. Rev.*, **133**, 3176–3201, doi:[10.1175/MWR3024.1](https://doi.org/10.1175/MWR3024.1).